Feature Contributions to ECG-based Heart-Failure Detection: Deep Learning vs. Statistical Analysis

Agnese Sbrollini[1], Chiara Leoni[1], Marjolein C. de Jongh[2], Micaela Morettini[1], Laura Burattini[1], Cees A. Swenne[2]

[1]Università Politecnica delle Marche, Ancona, Italy
[2]Leiden University Medical Center, Leiden, The Netherlands

Introduction. Knowing the contributions of ECG features to a specific ECG diagnosis is essential for the clinical interpretability of the diagnostic algorithm. Such an analysis is commonly done in conventional statistics, but seldomly addressed in deep learning. Here, we compare the contributions of ECG features in a deep-learning algorithm for the detection of emerging heart failure (HF) with those obtained by conventional statistical analysis.

Methods. We analyzed 58 ECG pairs obtained in myocardial infarction (MI) patients. Baseline ECGs were recorded ≥6 months after MI, without signs or symptoms of HF. Follow-up ECGs were recorded ≥12 months after MI (controls) or because of new HF-related complaints (cases). ECGs were characterized by 42 features.
A deep-learning neural network with 42 inputs and case/control outputs was created by the Repeated Structuring & Learning Procedure. For each patient, a feature ranking was constructed by analysis of the neural-network coefficients by a local-interpretable model-agnostic explanatory algorithm. The feature relevances (FR) for the study group were computed as 42 weighted averages (by ranking) of the percentage of patients presenting that specific feature in each of the ranking positions. Additionally, we computed 42 area-under-the-curve (AUC) values associated with conventional univariate analysis. To compare the deep-learning and univariate-based assessments of the importance of each of the 42 features, we computed Pearson's correlation coefficient (rho) between FR and AUC.

Results. Neural network training had a 99% classification performance and FR ranging from 4.47% (QRS complexity in follow-up ECG) to 0.32%. AUC ranged from 82% (absolute value of QRS-T spatial angle difference between baseline and follow-up ECGs) to 23%. Correlation analysis yielded no significant association between AUC and FR (rho=0.18, p>0.05).

Conclusion. Deep-learning and statistical-analysis generated feature contributions to heart-failure detection were discordant. Further studies will investigate which of the two approaches better reflects clinical interpretation.